



# UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE  
United States Patent and Trademark Office  
Address: COMMISSIONER FOR PATENTS  
P.O. Box 1450  
Alexandria, Virginia 22313-1450  
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/743,547	12/23/2003	Michal Louz-On	0026-0042	5793
44989	7590	05/12/2008		
HARRITY SNYDER, LLP 11350 Random Hills Road SUITE 600 FAIRFAX, VA 22030			EXAMINER TANG, KARIN C	
			ART UNIT 2151	PAPER NUMBER
			MAIL DATE 05/12/2008	DELIVERY MODE PAPER

**Please find below and/or attached an Office communication concerning this application or proceeding.**

The time period for reply, if any, is set in the attached communication.

# Office Action Summary

**Application No.**

10/743,547

**Applicant(s)**

LOUZ-ON, MICHAL

**Examiner**

KAREN C. TANG

**Art Unit**

2151

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --  
**Period for Reply**

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

**Status**

- 1) ☒ Responsive to communication(s) filed on 22 January 2008.
- 2a) ☒ This action is **FINAL**. 2b) ☐ This action is non-final.
- 3) ☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

**Disposition of Claims**

- 4) ☒ Claim(s) 1,3-5 and 7-29 is/are pending in the application.
- 4a) Of the above claim(s) \_\_\_\_\_ is/are withdrawn from consideration.
- 5) ☐ Claim(s) \_\_\_\_\_ is/are allowed.
- 6) ☒ Claim(s) 1,3-5 and 7-29 is/are rejected.
- 7) ☐ Claim(s) \_\_\_\_\_ is/are objected to.
- 8) ☐ Claim(s) \_\_\_\_\_ are subject to restriction and/or election requirement.

**Application Papers**

- 9) ☐ The specification is objected to by the Examiner.
- 10) ☐ The drawing(s) filed on \_\_\_\_\_ is/are: a) ☐ accepted or b) ☐ objected to by the Examiner.  
Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).  
Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).
- 11) ☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

**Priority under 35 U.S.C. § 119**

- 12) ☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).
- a) ☐ All b) ☐ Some \* c) ☐ None of:
1. ☐ Certified copies of the priority documents have been received.
  2. ☐ Certified copies of the priority documents have been received in Application No. \_\_\_\_\_.
  3. ☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

\* See the attached detailed Office action for a list of the certified copies not received.

**Attachment(s)**

- 1) ☒ Notice of References Cited (PTO-892)
- 2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
- 3) ☐ Information Disclosure Statement(s) (PTO-8508)  
Paper No(s)/Mail Date \_\_\_\_\_
- 4) ☐ Interview Summary (PTO-413)  
Paper No(s)/Mail Date \_\_\_\_\_
- 5) ☐ Notice of Informal Patent Application
- 6) ☐ Other: \_\_\_\_\_

- This action is responsive to the amendment and remarks file on 1/22/08.
- Claims 1, 3-5, 7-29 are presented for further examination.

### **DETAILED ACTION**

#### ***Response to Arguments***

Applicant's arguments with respect to claims 1, 3-5, 7-29 have been considered but are moot in view of the new ground(s) of rejection.

Applicant argues that Galai did not disclose locating the session identifiers in the set of URLs

Examiner disagrees.

Applicant's arguments fail to comply with 37 CFR 1.111(b) because they amount to a general allegation that the claims define a patentable invention without specifically pointing out how the language of the claims patentably distinguishes them from the references.

Galai discloses locating the session identifiers because its system contains capability to detect and identify a particular element or field or attributes (refer to 0067). Further, Galai's system must have capability to "identify" the session identifier (parameter) in order to remove the session identifier in the URL (refer to 0005, system/software are able to identify the parameter/session identifier).

#### ***Claim Rejections - 35 USC § 103***

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

(a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person

having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negated by the manner in which the invention was made.

Claims 1, 5, 7, 9, 10-13, 20-23 and 25-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over Galai (US 2004/0177015) in view of Bary (US 20040158429).

1. Referring to Claims 1, Galai discloses a method comprising: a set of uniform resource locators (URLs) from at least one document (refer to 0003); analyzing the extracted set of URLs extracted from the at least one document (content, being extracted from website, refer to 0011, 0015) to determine those in the set of URLs that contain session identifiers (detect and extract particular element, refer to 0019, 0022 and 0067) by locating the session identifiers in the set of URLs extracted as sub-strings that occur in multiple URL of a web site (refer to 0013, 0023, 0005 and 0067. It is obvious that in order to remove the session identifier, the system need to identify and locate where the session identifier is. The session identifier must be identified in order to be removed, see 0069. The act of identifying the session identifier is specifically indicated and also supported by Bary, par 0184, 0196, and 0205); generating a clean set of URLs from the set of URLs extracted form the at least one document by removing the session identifiers (refer to 0028); determining when at least one second URL has already been crawled based, at least in part, on a comparison of the second URL to the clean set of URLs (refer to 0020); The system of Galai's has already indicated the fact the system software will ignore any substring that appear to be session identify which is not part of URI (refer to 0005), although Galai did not specifically point out how the system is able to identify the session identifier, but

the system is able to remove the session identifier (parameter) out of the URL (refer to 0069). The act alone (removing the session identifier) shows that it must be obvious the system has capability to identify the session identifier in order to remove the session identifier. Bary, demonstrate the obviousness by disclosing the steps of how to identify the session identifier (refer to 0184, 0196). By demonstrate the obviousness and modifying the system of Galai, it would enable the user to structure a better website in order for the web host to have a better presence in the search engine.

2. Referring to Claims 10, 20 and 25, Galai discloses a method comprising: receiving a set of uniform resource locator (URLs, refer to 0020); analyzing the set of URL for sub-strings that are structured in a manner consistent with session identifiers (refer to 0015 and 0067, 0072); and further analyzing the set of URLs to identify of the sub-strings as corresponding to session identifiers based on multiple occurrence of a sub-string in the set of URLs (refer 0023 and 0067).

3. Referring to Claim 7, the method of claim 1, Galai discloses wherein the analyzing the set of URLs extracted from the at least one document further includes:

locating the session identifier in the extracted set of URLs as sub-strings that contain characters consistent with a session identifier (refer to 0020 and 0005, Barry also supported the act of locating the identifier, by extracting the characters that is consistent with a session identifier, refer to 0184 and 0196).

The system of Galai's has already indicated the fact the system software will ignore any substring that appear to be session identify which is not part of URI (refer to 0005), although

Galai did not specifically point out how the system is able to identify the session identifier, but the system is able to remove the session identifier (parameter) out of the URL (refer to 0069). The act alone (removing the session identifier) shows that it must be obvious the system has capability to identify the session identifier in order to remove the session identifier. Bary, demonstrate the obviousness by disclosing the steps of how to identify the session identifier (refer to 0184, 0196). By demonstrate the obviousness and modifying the system of Galai, it would enable the user to structure a better website in order for the web host to have a better presence in the search engine.

4. Referring to Claim 9, the method of claim 1, Galai discloses storing information based on the clean set of URLs for use in later determining whether additional URLs have already been extracted (extracted the document, and is stored temporally, refer to 0020); and storing the set of URLs extracted from the at least one document, including embedded session identifiers, for use in later accessing the set of URLs extracted from the at least one document (refer to 0184, 0196, and 0205).

5. Referring to Claims 11, 21 and 26, the device of claim 20, Galai discloses wherein the set of URLs are extracted from a web document associated with a web host (refer to 0013, 0070 and 0002).

6. Referring to Claims 12, 22 and 27, the device of claim 20, Galai discloses wherein the set of URLs are extracted from multiple documents associated with a single web host (refer to 0070).
7. Referring to Claims 13, 23 and 28, the device of claim 20, Galai discloses means for removing the identified session identifiers from the set of URLs (refer to 0069); and means for storing the set of URLs with the removed session identifiers as a clean set of URLs (refer to 0075).
8. Referring to Claim 5, Galai disclosed “wherein the session identifiers are determined as including sub-strings from the set of URLs that do not reference content.”(parameter/session identifier as visual layout characteristics, which does not references the content, refer to 0071)

Claims 3, 8, 15, 16, 17, 18 and 19 are rejected under 35 U.S.C. 103(a) as being unpatentable over Galai (US 2004/0177015) in view of Bary (US 20040158429) in further view of Applicant Admitted Prior Art hereinafter AAPA (Background Invention of applicant's specification, par 0002-0008)

9. Referring to Claim 3, the method of claim 1, Galai discloses wherein the at least one document is a web document from a web site (refer to 0061).

Art Unit: 2100

Although Galai and Bary disclosed the invention substantially as claim, Galai and Bary did not specifically indicate the web document is downloaded by the software.

AAPA disclosed web document is downloaded by the software (refer to 0004)

Hence, providing features specify by AAPA would let user to implement in order to detect and index the web page by the software and enable the software to analyze, parse, and index the dynamic web pages. .

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by AAPA.

10. Referring to Claim 19, Galai disclosed “wherein the session identifiers are determined as including sub-strings from the set of URLs that do not reference content.”(parameter/session identifier as visual layout characteristics, which does not references the content, refer to 0071)

11. Referring to Claim 15, Galai discloses at least one fetch bot (0009) configured to download content on network from locations specified by uniform resource locators/URLs (refer to 0020); a content manager configured to extract URLs from the content (refer to 0011), and identify sessions identifiers from URLs extracted from the content based (refer to 0020), at least in part, on multiple occurrences of the session identifiers from a single web sites (refer 0023); and a URL manager configured to store clean versions of the URLS (0075) extracted from the content in which the session identifiers are removed from the URLs extracted from the content (refer to 0019, 0028).



The system of Galai's has already indicated the fact the system software will ignore any substring that appear to be session identify which is not part of URI (refer to 0005), although Galai did not specifically point out how the system is able to identify the session identifier, but the system is able to remove the session identifier (parameter) out of the URL (refer to 0069). The act alone (removing the session identifier) shows that it must be obvious the system has capability to identify the session identifier in order to remove the session identifier. Bary, demonstrate the obviousness by disclosing the steps of how to identify the session identifier (refer to 0184, 0196). By demonstrate the obviousness and modifying the system of Galai, it would enable the user to structure a better website in order for the web host to have a better presence in the search engine.

Although Galai and Bary disclosed the invention substantially as claim, Galai and Bary did not specifically indicate the web document is downloaded by the software.

AAPA disclosed web document is downloaded by the software (refer to 0004)

Hence, providing features specify by AAPA would let user to implement in order to detect and index the web page by the software and enable the software to analyze, parse, and index the dynamic web pages. .

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by AAPA.

12. Referring to Claim 16, the device of claim 15, Galai discloses wherein the content manager is further configured to identify the session identifiers based on locating sub-strings, within the URLs extract from the content (refer 0023), the contain characters consistent with

session identifiers (refer to 0013, 0023, 0005 and 0067. It is obvious that in order to remove the session identifier, the system need to identify and locate where the session identifier is. The session identifier must be identified in order to be removed, see 0069. The act of identifying the session identifier is specifically indicated and also supported by Bary, par 0184, 0196, and 0205). The system of Galai's has already indicated the fact the system software will ignore any substrings that appear to be session identifier which is not part of URI (refer to 0005), although Galai did not specifically point out how the system is able to identify the session identifier, but the system is able to remove the session identifier (parameter) out of the URL (refer to 0069). The act alone (removing the session identifier) shows that it must be obvious the system has capability to identify the session identifier in order to remove the session identifier. Bary, demonstrate the obviousness by disclosing the steps of how to identify the session identifier (refer to 0184, 0196). By demonstrate the obviousness and modifying the system of Galai, it would enable the user to structure a better website in order for the web host to have a better presence in the search engine.

Although Galai and Bary disclosed the invention substantially as claim, Galai and Bary did not specifically indicate the web document is downloaded by the software.

AAPA disclosed web document is downloaded by the software (refer to 0004)

Hence, providing features specify by AAPA would let user to implement in order to detect and index the web page by the software and enable the software to analyze, parse, and index the dynamic web pages. .

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by AAPA.

13. Referring to Claim 17, the device of claim 15, Galai discloses further comprising: a database configured to store the content (refer to 0005).

Although Galai and Bary disclosed the invention substantially as claim, Galai and Bary did not specifically indicate to "download" the content.

AAPA disclosed web document is downloaded by the software (refer to 0004)

Hence, providing features specify by AAPA would let user to implement in order to detect and index the web page by the software and enable the software to analyze, parse, and index the dynamic web pages. .

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by AAPA.

14. Referring to Claim 8, the method of claim 1, Galai discloses further comprising: extract content from the particular URL when the particular URL is determined to not already have been crawled (refer to 0019). ).

Although Galai and Bary disclosed the invention substantially as claim, Galai and Bary did not specifically indicate to "download" the content.

AAPA disclosed the content is being downloaded (refer to 0004)

Hence, providing features specify by AAPA would let user to implement in order to detect and index the web page by the software and enable the software to analyze, parse, and index the dynamic web pages. .

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the system of Galai and Bary by including the features disclosed by AAPA.

15. Referring to Claim 18, the device of claim 15, Galai discloses wherein the URL manager is further configured to determine when additional URL have previously been stored by comparing clean version of the additional URLs to the stored clean versions of the URLs extracted from the content (refer to 0019, 0028).

Although Galai and Bary disclosed the invention substantially as claim, Galai and Bary did not specifically indicate to "download" the content.

AAPA disclosed the content is being downloaded (refer to 0004).

Hence, providing features specify by AAPA would let user to implement in order to detect and index the web page by the software and enable the software to analyze, parse, and index the dynamic web pages. .

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by AAPA.

Claims 4, 14, 24 and 29 are rejected under 35 U.S.C. 103(a) as being unpatentable over Galai (US 2004/0177015) in view of Bary (US 20040158429) in further view of Najork (US 6,952,730).

16. Referring to Claim 4, the method of claim 1, although Galai disclosed the invention substantially as claimed, Galai and Bary are silent regarding "wherein the comparison of the

second URL to the clean set of URL is based on a comparison of a fingerprint value calculated for each of the URLs in the clean set of URLs.”

Najork, in an analogous art disclosed “wherein the comparison of the second URL to the clean set of URL is based on a comparison of a fingerprint value calculated for each of the URLs in the clean set of URLs.” (refer to Col 9, Lines 4-17).

Hence, providing features disclosed by Najork, would be desirable for a user to implement to provide an efficient data structures that keep in tracks of downloaded document due to crawling the web pages.

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by Najork.

17. Referring to Claims 14, 24 and 29, Galai further discloses a method of claim 13, although Galai and Bary disclosed the invention substantially as claimed, Galai and Bary are silent regarding “adding a generated session identifier to URLs in the clean set of URLs when the URL are to be used to access a web document.”

Najork, in an analogous art disclosed, “adding a generated session identifier to URLs in the clean set of URLs when the URL are to be used to access a web document.” (refer to Col 6, Lines 55-67).

Hence, providing features disclosed by Najork, would be desirable for a user to implement to provide an efficient data structures that keep in tracks of downloaded document due to crawling the web pages.

Therefore, at the time of the invention, it would have been obvious to one of ordinary skill in the art to modify the systems of Galai and Bary by including the features disclosed by Najork.

### *Conclusion*

**Examiner's Notes:** Examiner has cited particular columns and line numbers in the references applied to the claims above for the convenience of the applicant. Although the specified citations are representative of the teachings of the art and are applied to specific limitations within the individual claim, other passages and figures may apply as well. It is respectfully requested from the applicant in preparing responses, to fully consider the references in entirety as potentially teaching all or part of the claimed invention, as well as the context of the passage as taught by the prior art or disclosed by the Examiner. In the case of amending the claimed invention, Applicant is respectfully requested to indicate the portion(s) of the specification which dictate(s) the structure relied on for proper interpretation and also to verify and ascertain the metes and bounds of the claimed invention.

Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37

CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Karen C. Tang whose telephone number is (571)272-3116. The examiner can normally be reached on M-F 7 - 3.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, John Follansbee can be reached on (571)272-3964. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see <http://pair-direct.uspto.gov>. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).

/K. C. T./  
Examiner, Art Unit 2151

/John Follansbee/  
Supervisory Patent Examiner, Art Unit 2151